DOCUMENT RESUME

ED 050 173 TM 000 570

AUTHOR Campbell, Paul B.; Beers, Joan S

TITLE Definition and Measurement in the Affective Domain: Appreciation of Human Accomplishments.

INSTITUTION Pennsylvania State Dept. of Education, Harrisburg.
PUB DATE Feb 71
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Affective Behavior, Factor Analysis, *Grade 5, *Grade 11, *Measurement Techniques, Pictorial Stimuli, Reliability, *Research Tools, Test Construction, Visual Measures
IDENTIFIERS Pennsylvania Inventory of Cultural Appreciations, *Pennsylvania Plan, PICA, Things People Do Inventory, TPD

ABSTRACT

The first three levels of the taxonomy in the affective domain guided the development of two inventories--the Pennsylvania Inventory of Cultural Appreciations (PICA) for 11th graders and Things People Do (TPD) for 5th graders--to measure appreciation of human accomplishments in seven areas: politics, sciences, sports, literature, visual arts, music, and theatre. Alpha coefficient for the PICA is .92, and for the TPD, .75. Photographs are used in the TPD. With a sample of 3,000 at each grade level, analyses of the instruments included comparison of means, analysis of correlations, factor analysis and analysis of cumulative structure. The hypothesized hierarchical structure is supported by the TPD data but not by the PICA data. Many inferences and implications are discussed. (Author/GS)

Definition and Measurement in the Affective Domain:

Appreciation of Human Accomplishments

Paul B. Campbell and Joan S. Beers
Pennsylvania Department of Education

Objective

One of the stated goals of education in Pennsylvania is that pupils should show appreciation for human accomplishments in the natural sciences, the social sciences, the humanities and the arts. The Bureau of Educational Quality Assessment in the Pennsylvania Department of Education was given the assignment to measure the degree to which pupils are showing these appreciations. In order to assess the level of pupil achievements, measuring inventories were needed. A search for existing inventories proved fruitless. Instrumentation in the area of cultural appreciations reflects the lack of attention traditionally paid this area in the curriculum. A further limitation was imposed by the corresponding lack of attention traditionally paid the affective domain in the measurement field.

The objective of this paper is to describe the process of defining and measuring in the affective domain through the development of two inventories -- Pennsylvania Inventory of Cultural Appreciations (PICA) for 11th graders and Things People Do (TPD) for 5th graders.

Method

PICA was developed first. Human accomplishments were divided into seven categories:

1. Politics
2. Sciences
3. Sports
4. Literature
5. Visual Arts
6. Music
7. Theater

The first three levels of the Taxonomy of Educational Objectives: Affective Domain (Krathwohl, Bloom, Masia, 1956) were used to guide the construction of items. Items were developed to measure in each of the seven categories at each of the first three levels of the taxonomy.

Level 1: Receiving (Attending) At this level the learner is expected to be willing to receive or attend to a certain stimulus rather than to avoid it.

Sample Items -- If you had the opportunity, would you like to: Visit an art museum? See a soccer game? Attend an opera?

Level 2: Responding At this level the learner is expected to move from merely receiving a stimulus to actively attending. Also at this level not only does the learner actively attend but begins to gain satisfaction in responding. Music becomes an emotional involvement. Politics become a zealous pursuit. Reading becomes a vicarious experience.

Sample Items -- (1) Poetry is best when I can read it aloud to myself. (2) Most works of art are too difficult to understand. (3) I would like to be able to vote in a state or national election.

Level 3: Valuing At this level the learner is expected to display a certain behavior with sufficient consistency to show that he perceives the behavior to have worth. At this lowest level of valuing, commitment is still uncertain. The learner continues to reevaluate his beliefs and his position is still tentative.

Sample Items -- (1) Most scientists are interested only in machines, not in people. (2) Most people in public life are dedicated to helping the public. (3) Tax money should be used to support the theater.

In early September of 1968, seventy-nine items were tried out with five 11th grade groups: an "average" ability group from an urban school (N = 23); the "highest" and "lowest" ability groups from a rural school (N = 59); and an "honors" group and a "general" group from a suburban school (N = 54). The total sample was 136 11th grade students. Post-test individual and group interviews were held to collect student opinions about the clarity or ambiguity of the items, the appropriateness of the response options and any other reactions.

As a result of the tryout, seemingly ambiguous items were rewritten, imbalance was corrected with an equal number of items for each of the seven categories and nondiscriminating items were eliminated or revised. The format was refined and the first 21 items were designated "knowledge" rather than affective items. The resulting PICA contains 77 items:

21 Knowledge items
21 Receiving items
21 Responding items
14 Valuing items

PICA was then administered to a group of urban 5th grade pupils to test its applicability at the lower grade level. The results indicated that the vocabulary and the level of comprehension required were too advanced for many 5th grade pupils. Things People Do was then constructed for 5th grade pupils.

TPD attempts to measure pupil appreciations in the same seven categories according to the first three levels of the taxonomy. Since reading level is of utmost concern for 5th grade pupils, a group of photographs is used to depict each of the seven categories. The inventory asks the pupil to consider a group of four pictures and then to answer three questions about the pictures. For example, in the music category the pictures depict a symphony orchestra, ballet dancers, a rock group and a vocalist. The pupil is asked:

1. Would you like to see or hear one or more of these kinds of musical activities? (Receiving)

2. Would you like to take part in one or more of these kinds of musical activities? (Responding)

3. Do you think music is very important, somewhat important or not important? (Valuing)

TPD contains 21 items:

7 Receiving items
7 Responding items
7 Valuing items

Sample

In the fall of 1969, the inventories were administered to 20,000 5th graders and 20,000 11th graders as part of a statewide assessment battery. The results from these large samples were used to develop norms for both inventories.

Subsamples of approximately 3,000 at each grade level were drawn in order to analyze the technical properties of the inventories. The subsamples are representative of pupils from districts of varying degrees of wealth and proportionately representative of pupils from schools of varying sizes.

Analysis and Results

Reliability. The reliability estimates of inventories such as PICA and TPD present several problems for which only partial solutions exist. A test-retest coefficient would be most appropriate for this analysis if the necessary conditions could be met (i.e., minimal practice effect and/or the existence of parallel forms and the availability of the sample for retesting). These conditions could not be met at the present stage of inventory development. Overall estimates of coefficient alpha (Cronbach, 1951) were, therefore, computed for both TPD and PICA. It is recognized that this statistic may provide an unprecise estimate of reliability if hierarchiality exists. The coefficients can be considered a conservative estimate of reliability under the assumption that the strata are fixed rather than random. The strata were defined as fixed (Rabinowitz and Eikeland, 1964). The obtained alpha values for the TPD and PICA are .79 and .92, respectively.

The inventories were designed to represent a hierarchy of affective behaviors associated with human accomplishments in the sciences, arts and humanities. The empirical confirmation of the existence of this hierarchy was approached in several ways.

Comparison of Means. The first analysis considered differences among strata means. The concept of level of involvement is used to serve the same function in the affective domain as the concept of level of complexity serves in the cognitive domain (Kropp and Stoker, 1966). The theory of an affective hierarchy requires that valuing represent a higher level of involvement than responding, which in turn is higher in involvement than receiving. It is, therefore, hypothesized that the mean score for receiving will be highest, for responding less high and for valuing least high. Evidence of such ordering is present in the first two levels of TPD, but is reversed in PICA (See Table 1).

Analysis of Correlations. An analysis of the inter-item correlation matrix analogous to the Guttman Radex Technique (Guttman, 1964) was undertaken with TPD. If the hierarchy exists, then it is expected that inter-item correlations will be greatest among items within levels, next in magnitude between items in adjacent levels and least in magnitude between items in more remote levels. Although evidence for the hierarchy exists in the TPD data, the correlations do not completely fit the hypothesized order. Two phenomena are present. First, the inter-item correlations within levels are significantly higher than the inter-item correlations across either adjacent or remote levels as expected. However, the order of magnitude of the mean inter-item correlations between adjacent and between remote levels are reversed from the hypothesized order (See Table 2). Second, there appears to be more item variance common within categories of activity across levels than within levels across categories of activities. The mean inter-item correlation within activity categories is .47. The disparate nature of the activities represented probably accounts for this phenomenon and also serves to suppress somewhat the emergence of the hierarchical structure.

This situation is further reinforced when the pattern of responses to individual items within each category is examined. In two of the seven categories (politics and science) the mean item response for valuing is higher than for the first two levels, which is a reversal of the hypothesized direction. In the remaining five categories (music, art, literature, theater and sports) the means follow the inter-item correlation pattern predicted by the hierarchy (See Table 3).

Several possible explanations are considered for the deviations from the pattern. The hierarchy may overlap at the responding and valuing levels. A person may need to value an activity before he can respond, thereby confirming the overall correlational pattern between these levels. Furthermore, the items may not have elicited in two categories the appropriate responses from a substantial number of students. A more likely explanation is that politics and science deviated so sharply from the pattern that its existence in the correlation matrix is submerged.

A similar analysis was conducted with PICA. Because the patterns of mean item responses suggested an interaction between categories of activity and levels of the hierarchy, a modification of the analysis procedure was undertaken. In addition to the sets of within levels, adjacent levels and remote levels inter-item correlations, the set of inter-item correlations which were specific to each activity category was removed and treated as a separate category. The hypothesized order of magnitude of the average correlation for each of these sets was as follows:

$r_{\text{within activities}} > r_{\text{within levels}} > r_{\text{adjacent levels}} > r_{\text{remote levels}}$

As in the case of the TPD analysis, there is a significant difference among the mean correlations. The order in this instance is exactly congruent with the hypothesized order (See Table 4). It would appear then that if this

pattern holds up in subsequent data collections, the existence of an affective hierarchy in the area of appreciation of human accomplishments is supported. The definition, however, of the levels of the hierarchy by the measuring instrument presents some problems. These will be discussed in a later section.

Factor Analysis. The inventories were factor analyzed as part of a battery and as separate inventories. The results support TPD as a separate inventory apart from inventories to measure citizenship, creativity, inter-personal relationships and interest in school. The factor analysis of TPD as a separate inventory produced a principal components solution with a general factor upon which each item has a factor loading of at least .40. A varimax rotation produced four factors which accounted for 54% of the explained variance. Factor one included items from art, theater and music. Factor two included items from science and literature. Factors three and four are defined by items from politics and sports in that order. The three levels of the taxonomy - receiving, responding and valuing - are not discernible in the factor structure.

The factor analysis for the 11th grade instrument, PICA, also supported its use as a separate measure from others included in the total battery. All items have a loading of at least .53 in the general factor from the principal components solution. Three factors were defined by the varimax rotation. They explained a total of 60% of the variance. Factor one accounted for theater, art and valuing items. Factor two included sports, politics and science items. Factor three was defined by music and a general set of receiving items. The hypothesized literature items were represented among all three factors. Two levels of the affective hierarchy seem to be emerging as separate factors in this analysis. For more detailed information on this aspect of the study see Beers, 1971.

Analysis of Cumulative Structure. Additional analyses were designed to test the hypothesized property of cumulativeness. The theory of hierarchiality requires that in order to value one must respond and in order to respond one must receive. Corollary to this hypothesis is the expectation that the highest level of involvement, valuing, will be less frequently represented in the population than responding, which, in turn, will be of a lesser frequency than receiving. To test these hypotheses the patterns of response to the items on the two inventories (TPD and PICA) were organized into sets according to the degree of consistency between the pattern and the hierarchy. Strong consistent patterns were those which showed a high level of receiving, a lower level of responding and the lowest level of valuing. This set is consistent with the expected order of frequencies of the affective levels in the population. The weak consistent set included those patterns of response which were not contradictory in the sense that a person valued without being willing to receive. A pattern which indicated that a student was moderately willing to receive, to respond and to value would be assigned to this set. The inconsistent set of patterns included those which showed high levels of valuing, moderate levels of responding and an unwillingness to receive. In order to exhaust the possibilities of response patterns to the instrument, all items were assigned to a triad consisting of receiving, responding and valuing items. The directionality of the triad was examined by considering the difference between the mean of the first pair of the triad and the mean of the second pair. If the first mean was larger than the second, then the item was considered to be supportive of the hierarchy and assigned to the strong set. If the means of the two pairs were equal, then the triad was considered consistent and assigned to the weak set of response patterns. If the mean of the second pair was larger than the mean of the first pair, then the triad

was considered inconsistent with the hierarchy and was assigned to the inconsistent set. The response patterns were defined within each activity category and a computer program was devised to sort the population of subjects into each of the three mutually exclusive sets. It was predicted that if the affective domain hierarchy was working as defined by these measures in this population, the largest frequencies would occur in the strong set, the next lowest in the weak set, and the lowest frequencies in the inconsistent set. The data for TPD show that for sports, visual arts and theater, the predicted patterns hold. For literature and music the sum of the frequencies in the strong and weak categories is significantly greater than those in the inconsistent categories. For politics and science the weak category contains the greatest frequency with the inconsistent category containing slightly greater frequencies than the strong categories. In general, this analysis is judged to support somewhat the existence of the affective hierarchy as defined with a possible explanation of the inconsistency in politics and science to be found in the lack of opportunities to participate in these activities by 5th grade students (See Table 5).

The case of the PICA is quite different. In sports, politics, music and literature the greatest frequencies are found in the inconsistent category, the next greatest in the weak category, and the lowest frequencies in the strong set. In natural science and visual arts the greatest frequencies are found in the weak set, but the inconsistent set also has greater frequencies than the strong set. Only in theater movies does the pattern fall in the predicted direction (See Table 5). These data together with the correlation analysis suggest that an affective hierarchy does indeed exist at both the 5th and 11th grade levels as represented by the TPD and PICA, but a redefinition of the levels in terms of their item representation and an operational definition of the concepts of responding and valuing needs to be undertaken.

This conclusion is further borne out by an additional analysis. A panel of four judges independently assigned each item on the PICA to the three levels - receive, respond and value. There was only one instance of disagreement on the 21 receiving items. In contrast there were only 6 out of 21 responding items upon which there was agreement and 11 out of 14 valuing items. A major problem appears to be the definition of responding items as distinct from valuing items.

Summary

In summary the researchers conclude that the development and analysis of TPD and PICA demonstrate several things.

> Reliable measurements in the hard-to-measure affective area can be constructed.
>
> Photographs can successfully be used to elicit responses from elementary pupils.
>
> The levels of the taxonomy can form a theoretical basis for developing inventories in the affective domain.
>
> The levels of the taxonomy emerge as distinct when several kinds of statistical techniques are applied.
>
> The hypothesized hierarchical structure of the taxonomy appears to exist with TPD for 5th graders. A hierarchical structure appears also to exist with PICA for 11th graders, but not in the hypothesized direction.

The work on defining operationally the hierarchical structure at the 11th grade must be considerably extended. Continuous development and tryout of both instruments with new samples are underway.

Table 1

Mean Item Responses

Inventories	Levels		
	Receiving	Responding	Valuing
TPD Grade 5 N =	2.66	2.37	2.37
PICA Grade 11 N =	2.17	2.28	2.46

Table 2

Summary of Analysis of Variance for Inter-Item Association - TPD

Source of variation	Sum of Squares	Degree of Freedom	Mean Square
Between categories	$SS_{categories} = 0.1819$	2.0000	$MS_{categories} = 0.0909$
Experimental error	$SS_{error} = 0.6449$	123.0000	$MS_{error} = 0.0052$
Total	$SS_{total} = 0.8268$	125.0000	$F = 17.4807$

	Within Levels	Adjacent Levels	Remote Levels
$\bar{X}_r$	0.3476	0.2633	0.2971

Table 3

Appreciation of Human Accomplishments
Average Item Response by Activity Categories

TPD Grade 5	Activity Categories						
	Politics	Science	Sports	Literature	Visual Arts	Music	Theater
Receiving	2.28	2.76	2.88	2.76	2.75	2.46	2.74
Responding	2.01	2.34	2.68	2.65	2.50	2.21	2.23
Valuing	2.59	2.83	2.29	2.53	2.24	2.17	1.99

FICA Grade 11	Sports	Natural Science	Politics	Theater, Movies	Music, Dance	Visual Arts	Literature
Receiving	2.29	2.20	1.92	2.72	1.70	2.26	2.10
Responding	2.52	2.10	2.14	2.63	2.28	2.05	2.26
Valuing	2.86	2.49	2.19	2.15	2.46	2.61	2.49

Table 4

Summary of Analysis of Variance for Inter-Item Association - PICA

Source of variation	Sum of Squares	Degree of Freedom	Mean Square
Between categories	$SS_{categories}$ = 1.0571	3.0000	$MS_{categories}$ = 0.3523
Experimental error	SS_{error} = 11.6348	1538.0000	MS_{error} = 0.0075
Total	SS_{total} = 12.6919	1541.0000	F = 46.9733

	Within Activities	Within Levels	Adjacent Levels	Remote Levels
$\bar{X}_r$	0.5569	0.5020	0.4826	0.4677

Table 5

Item Pattern Frequencies

TPD Grade 5	Activity Categories						
	Politics	Science	Sports	Literature	Visual Arts	Music	Theater
Strong	343	267	1460	742	1385	1089	1698
Weak	1340	2008	1067	1628	1061	1176	807
Inconsistent	927	335	83	240	164	345	105

PICA Grade 11	Sports	Natural Science	Politics	Theater, Movies	Music, Dance	Visual Arts	Literature
Strong	94	249	440	1462	146	273	226
Weak	1035	1315	933	843	759	1185	1102
Inconsistent	1324	889	1080	148	1548	995	1125

References

Beers, J. S., Educational Quality Assessment, Phase II Findings, Section 4, The Ten Goals of Quality Education, 1971

Cronbach, L. J., Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Guttman L., A new approach to factor analysis: the radex. In Lazarsfeld, P. (ed.), Mathematical Thinking in the Social Sciences. Glencoe, Illinois: The Free Press, 1964.

Krathwohl, D. R., Bloom, B. S., and Masia, B. B., Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Company, 1956.

Kropp, R. P., Stoker, H. W., and Bashaw, U. L., The Construction and Validation of Tests of the Cognitive Processes as Described in the Taxonomy of Educational Objectives. Florida State University: Institute of Human Learning and Department of Educational Research and Testing, 1966.

Rabinowitz, W. and Eikeland, H. M., Estimating the reliability of tests with clustered items. Pedagogisk Forskning, 1964, 85-106.